第二章

最小二乘估计（二）

Outline

1. 渐近理论基础
2. 最小二乘估计的大样本性质
3. 大样本统计推断
4. Resampling Methods: Jackknife and Bootstrap

渐近理论基础

Various Modes of Convergence

Definition (Convergence in Probability)

A sequence of random scalars $\{z_n\}$ converges in probability to a constant α if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{Prob}(|z_n - \alpha| > \varepsilon) = 0.$$

The constant α is called the probability limit of z_n and is written as $\operatorname*{plim}_{n\to\infty} z_n = \alpha$ or $z_n \underset{p}{\to} \alpha$.

A sequence of K-dimensional random vectors $\{\mathbf{z}_n\}$ converges in probability to a K-dimensional vector of constants $\boldsymbol{\alpha}$ if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{Prob}(|z_{nk} - \alpha_k| > \varepsilon) = 0 \quad \text{for all } k = 1, 2, \ldots, K,$$

where z_{nk} is the k-th element of $\mathbf{z}_n$ and α_k the k-th element of $\boldsymbol{\alpha}$.

Definition (Almost Sure Convergence)

A sequence of random scalars $\{z_n\}$ converges almost surely to a constant α if

$$\text{Prob}\left(\lim_{n\to\infty} z_n = \alpha\right) = 1.$$

We write this as $z_n \underset{a.s.}{\rightarrow} \alpha$. The extension to random vectors is analogous to that for convergence in probability.

Definition (Convergence in Mean Square)

A sequence of random scalars $\{z_n\}$ converges in mean square (or in quadratic mean) to a constant α (written as $z_n \underset{m.s.}{\rightarrow} \alpha$) if

$$\lim_{n\to\infty} \mathbb{E}\left[(z_n - \alpha)^2\right] = 0.$$

Convergence to a Random Variable

We say that a sequence of K-dimensional random variables $\{\mathbf{z}_n\}$ converges in probability to a K-dimensional random variable $\mathbf{z}$ and write $\mathbf{z}_n \underset{p}{\to} \mathbf{z}$ if $\{\mathbf{z}_n - \mathbf{z}\}$ converges in probability to $\mathbf{0}$:

$$\text{“}\mathbf{z}_n \underset{p}{\to} \mathbf{z}\text{”} \quad \text{is the same as} \quad \text{“}\mathbf{z}_n - \mathbf{z} \underset{p}{\to} \mathbf{0}.\text{”}$$

Similarly,

$$\text{“}\mathbf{z}_n \underset{a.s.}{\to} \mathbf{z}\text{”} \quad \text{is the same as} \quad \text{“}\mathbf{z}_n - \mathbf{z} \underset{a.s.}{\to} \mathbf{0}.\text{”}$$

$$\text{“}\mathbf{z}_n \underset{m.s.}{\to} \mathbf{z}\text{”} \quad \text{is the same as} \quad \text{“}\mathbf{z}_n - \mathbf{z} \underset{m.s.}{\to} \mathbf{0}.\text{”}$$

Definition (Convergence in Distribution)

Let $\{z_n\}$ be a sequence of random scalars and F_n be the cumulative distribution function (c.d.f) of z_n. We say that $\{z_n\}$ converges in distribution to a random scalar z if the c.d.f. F_n of z_n converges to the c.d.f. F of z at every continuity point of F. We write $z_n \underset{d}{\rightarrow} z$ or $z_n \underset{L}{\rightarrow} z$ and call F the asymptotic (or limit or limiting) distribution of z_n.

Theorem (Multivariate Convergence in Distribution Theorem)

Let $\{\mathbf{z}_n\}$ be a sequence of K-dimensional random vectors. Then:

$\mathbf{z}_n \underset{d}{\rightarrow} \mathbf{z} \;\Leftrightarrow\; \boldsymbol{\lambda}'\mathbf{z}_n \underset{d}{\rightarrow} \boldsymbol{\lambda}'\mathbf{z}$ *for any K-dimensional vector of real numbers $\boldsymbol{\lambda}$.*

Remark: For convergence in distribution, element-by-element convergence does not necessarily mean convergence for the vector sequence.

Theorem (Relationship among the four modes of convergence)

(a) $\mathbf{z}_n \underset{m.s.}{\rightarrow} \boldsymbol{\alpha} \;\Rightarrow\; \mathbf{z}_n \underset{p}{\rightarrow} \boldsymbol{\alpha}, \quad \mathbf{z}_n \underset{m.s.}{\rightarrow} \mathbf{z} \;\Rightarrow\; \mathbf{z}_n \underset{p}{\rightarrow} \mathbf{z}.$

(b) $\mathbf{z}_n \underset{a.s.}{\rightarrow} \boldsymbol{\alpha} \;\Rightarrow\; \mathbf{z}_n \underset{p}{\rightarrow} \boldsymbol{\alpha}, \quad \mathbf{z}_n \underset{a.s.}{\rightarrow} \mathbf{z} \;\Rightarrow\; \mathbf{z}_n \underset{p}{\rightarrow} \mathbf{z}.$

(c) $\mathbf{z}_n \underset{p}{\rightarrow} \boldsymbol{\alpha} \;\Leftrightarrow\; \mathbf{z}_n \underset{d}{\rightarrow} \boldsymbol{\alpha}.$ *(That is, if the limiting random variable is a constant [a trivial random variable], convergence in distribution is the same as convergence in probability.)*

Viewing Estimators as Sequences of Random Variables

Let $\hat{\boldsymbol{\theta}}_n$ be an estimator of a parameter vector $\boldsymbol{\theta}$ based on a sample of size n. We say that an estimator $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}$ if $\hat{\boldsymbol{\theta}}_n \underset{p}{\rightarrow} \boldsymbol{\theta}$.

Continuous Mapping Theorem

Theorem (Continuous Mapping Theorem)

Let $\mathbf{a}(\cdot)$ *be a vector-valued function that does not depend on* n.

(a) *Suppose* $\mathbf{a}(\cdot)$ *is continuous at* $\boldsymbol{\alpha}$. *Then,* $\mathbf{z}_n \underset{p}{\to} \boldsymbol{\alpha} \ \Rightarrow \ \mathbf{a}(\mathbf{z}_n) \underset{p}{\to} \mathbf{a}(\boldsymbol{\alpha})$.
Stated differently,

$$\underset{n\to\infty}{\text{plim}}\ \mathbf{a}(\mathbf{z}_n) = \mathbf{a}(\underset{n\to\infty}{\text{plim}}\ \mathbf{z}_n).$$

(b) *Suppose* $\mathbf{a}(\cdot)$ *is continuous almost everywhere. Then,*

$$\mathbf{z}_n \underset{d}{\to} \mathbf{z} \ \Rightarrow \ \mathbf{a}(\mathbf{z}_n) \underset{d}{\to} \mathbf{a}(\mathbf{z}).$$

Slutsky's Theorem

Theorem (Slutsky's Theorem)

(a) "$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x},\ \mathbf{y}_n \underset{p}{\rightarrow} \boldsymbol{\alpha}$" $\Rightarrow$ "$\mathbf{x}_n + \mathbf{y}_n \underset{d}{\rightarrow} \mathbf{x} + \boldsymbol{\alpha}$."

(b) "$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x},\ \mathbf{y}_n \underset{p}{\rightarrow} \mathbf{0}$" $\Rightarrow$ "$\mathbf{y}_n'\mathbf{x}_n \underset{p}{\rightarrow} 0$."

(c) "$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x},\ \mathbf{A}_n \underset{p}{\rightarrow} \mathbf{A}$" $\Rightarrow$ "$\mathbf{A}_n\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{A}\mathbf{x}$." *In particular, if* $\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, *then* $\mathbf{A}_n\mathbf{x}_n \underset{d}{\rightarrow} N(\mathbf{0}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

(d) "$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x},\ \mathbf{A}_n \underset{p}{\rightarrow} \mathbf{A}$" $\Rightarrow$ "$\mathbf{x}_n'\mathbf{A}_n^{-1}\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$," *provided* $\mathbf{A}$ *is nonsingular.*

Delta Method

Theorem (Delta Method)

Suppose $\{\mathbf{x}_n\}$ *is a sequence of* K*-dimensional random vectors such that* $\mathbf{x}_n \underset{p}{\to} \boldsymbol{\beta}$ *and*

$$\sqrt{n}(\mathbf{x}_n - \boldsymbol{\beta}) \underset{d}{\to} \mathbf{z},$$

and suppose $\mathbf{a}(\cdot) : \mathbb{R}^K \to \mathbb{R}^r$ *has continuous first derivatives with* $\mathbf{A}(\boldsymbol{\beta})$ *denoting the* $r \times K$ *matrix of first derivatives evaluated at* $\boldsymbol{\beta}$*:*

$$\mathbf{A}(\boldsymbol{\beta}) \equiv \frac{\partial \mathbf{a}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

Then

$$\sqrt{n}[\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\boldsymbol{\beta})] \underset{d}{\to} \mathbf{A}(\boldsymbol{\beta})\mathbf{z}.$$

In particular:

$$\sqrt{n}(\mathbf{x}_n - \boldsymbol{\beta}) \underset{d}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}) \;\Rightarrow\; \sqrt{n}[\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\boldsymbol{\beta})] \underset{d}{\to} N(\mathbf{0}, \mathbf{A}(\boldsymbol{\beta})\boldsymbol{\Sigma}\mathbf{A}(\boldsymbol{\beta})').$$

Proof.

由中值定理，存在 $\mathbf{y}_n$ between $\mathbf{x}_n$ and $\boldsymbol{\beta}$ 使得：

$$\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\boldsymbol{\beta}) = \mathbf{A}(\mathbf{y}_n)(\mathbf{x}_n - \boldsymbol{\beta}).$$

等式两边同时乘以 $\sqrt{n}$,

$$\sqrt{n}[\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\boldsymbol{\beta})] = \mathbf{A}(\mathbf{y}_n)\sqrt{n}(\mathbf{x}_n - \boldsymbol{\beta}).$$

因为 $\mathbf{x}_n \underset{p}{\rightarrow} \boldsymbol{\beta}$，所以：$\mathbf{y}_n \underset{p}{\rightarrow} \boldsymbol{\beta}$. 再由连续映射定理可以得到：

$$\mathbf{A}(\mathbf{y}_n) \underset{p}{\rightarrow} \mathbf{A}(\boldsymbol{\beta}).$$

根据 Slutsky's Theorem 和 $\sqrt{n}(\mathbf{x}_n - \boldsymbol{\beta}) \underset{d}{\rightarrow} \mathbf{z}$ 易知定理结论成立。 □

大数定律和中心极限定理

Theorem (Law of Large Numbers, LLN)

- *Khintchine's weak LLN: Let $\{z_i\}$ be i.i.d. with a finite mean μ. Then*

$$\bar{z}_n \equiv \frac{1}{n}\sum_{i=1}^{n} z_i \underset{p}{\rightarrow} \mu.$$

- *Kolmogorov's strong LLN: Let $\{z_i\}$ be i.i.d. random variables. Then*

$$\bar{z}_n \underset{a.s.}{\rightarrow} \mu \quad \textit{iff} \quad \mathbb{E}|z_i| < \infty \textit{ and } \mathbb{E}(z_i) = \mu.$$

Theorem (Lindeberg-Levy Central Limit Theorem, CLT)

Let $\{\mathbf{z}_i\}$ be i.i.d. with $\mathbb{E}(\mathbf{z}_i) = \boldsymbol{\mu}$ and $\mathrm{Var}(\mathbf{z}_i) = \boldsymbol{\Sigma}$. Then

$$\sqrt{n}(\bar{\mathbf{z}}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{z}_i - \boldsymbol{\mu}) \underset{d}{\rightarrow} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

时间序列分析的基本概念

(Strictly) Stationarity

The process $\{\mathbf{z}_i\}$ $(i = 1, 2, \dots)$ is said to be (strictly) stationary if $(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_r})$ and $(\mathbf{z}_{i_1+k}, \dots, \mathbf{z}_{i_r+k})$ have the same joint distribution, for any $r \in \mathbb{N}$ and any $k \in \mathbb{Z}$.

Weakly Stationarity

The process $\{\mathbf{z}_i\}$ is said to be weakly stationary (or covariance stationary or second-order stationary) if:

(i) $\mathbb{E}(\mathbf{z}_i)$ does not depend on i, and

(ii) $\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_{i-j})$ exists, is finite, and depends only on j but not on i.

Ergodicity

A stationary process $\{z_i\}$ is said to be ergodic if, for any two bounded functions $f : \mathbb{R}^{k+1} \to \mathbb{R}$ and $g : \mathbb{R}^{l+1} \to \mathbb{R}$,

$$\lim_{n\to\infty} \left| \mathbb{E}[f(z_i, \ldots, z_{i+k}) g(z_{i+n}, \ldots, z_{i+n+l})] \right| = \left| \mathbb{E}[f(z_i, \ldots, z_{i+k})] \right| \left| \mathbb{E}[g(z_i, \ldots, z_{i+l})] \right|.$$

A stationary process that is ergodic will be called ergodic stationary.

Remark: Heuristically, a stationary process is ergodic if it is asymptotically independent.

Theorem (Ergodic Theorem)

Let $\{\mathbf{z}_i\}$ be a stationary and ergodic process with $\mathbb{E}(\mathbf{z}_i) = \boldsymbol{\mu}$. Then

$$\bar{\mathbf{z}}_n \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \underset{a.s.}{\rightarrow} \boldsymbol{\mu}.$$

Martingale

A vector process $\{\mathbf{z}_i\}$ is called a martingale if

$$\mathbb{E}(\mathbf{z}_i \mid \mathbf{z}_{i-1}, \ldots, \mathbf{z}_1) = \mathbf{z}_{i-1} \quad \text{for } i \geq 2.$$

Martingale Difference Sequence

A vector process $\{\mathbf{g}_i\}$ with $\mathbb{E}(\mathbf{g}_i) = \mathbf{0}$ is called a martingale difference sequence (m.d.s.) if the expectation conditional on its past values, too is zero:

$$\mathbb{E}(\mathbf{g}_i \mid \mathbf{g}_{i-1}, \ldots, \mathbf{g}_1) = \mathbf{0} \quad \text{for } i \geq 2.$$

A martingale difference sequence has no serial correlation.

Theorem (Ergodic Stationary Martingale Differences CLT)

Let $\{\mathbf{g}_i\}$ be a vector martingale difference sequence that is stationary and ergodic with $\mathbb{E}(\mathbf{g}_i\mathbf{g}_i') = \mathbf{\Sigma}$, and let $\bar{\mathbf{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i$. Then

$$\sqrt{n}\bar{\mathbf{g}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{g}_i \underset{d}{\rightarrow} N(\mathbf{0}, \mathbf{\Sigma}).$$

Remark:

- This CLT, being applicable not just to i.i.d. sequences but also to stationary martingale differences such as ARCH(1) processes, is more general than Lindeberg-Levy.
- Ergodic Theorem is an LLN for serially correlated processes. A central limit theorem for serially correlated processes can be found in Section 6.5 of [Hayashi, 2000].

最小二乘估计的大样本性质

大样本性质的推导基于以下关于数据生成过程（Data Generating Process, DGP）的假设：

1. Assumption 2.1 (linearity):

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, 2, \ldots, n)$$

   where $\mathbf{x}_i$ is a K-dimensional vector of regressors, $\boldsymbol{\beta}$ is a K-dimensional coefficient vector, and ε_i is the unobservable error term.

2. Assumption 2.2 (ergodic stationarity): The $(K+1)$-dimensional vector stochastic process $\{y_i, \mathbf{x}_i\}$ is jointly stationary and ergodic.

3. Assumption 2.3 (predetermined regressors): All the regressors are predetermined in the sense that they are orthogonal to the contemporaneous error term: $\mathbb{E}(x_{ik}\varepsilon_i) = 0$ for all i and k $(= 1, 2, \ldots, K)$. This can be written as $\mathbb{E}[\mathbf{x}_i \cdot (y_i - \mathbf{x}_i'\boldsymbol{\beta})] = \mathbf{0}$ or $\mathbb{E}(\mathbf{g}_i) = \mathbf{0}$ where $\mathbf{g}_i \equiv \mathbf{x}_i \cdot \varepsilon_i$.

4. Assumption 2.4 (rank condition): The $K \times K$ matrix $\mathbb{E}(\mathbf{x}_i\mathbf{x}_i')$ is nonsingular (and hence finite). We denote this matrix by $\mathbf{\Sigma_{xx}}$.
5. Assumption 2.5 ($\mathbf{g}_i$ is a martingale difference sequence with finite second moments): $\{\mathbf{g}_i\}$ is a martingale difference sequence. The $K \times K$ matrix of cross moments, $\mathbb{E}(\mathbf{g}_i\mathbf{g}_i')$, is nonsingular.

Remark

- Let $\mathbf{S} \equiv \mathrm{Avar}(\bar{\mathbf{g}})$, the variance of the asymptotic distribution of $\sqrt{n}\bar{\mathbf{g}}$. Then, $\mathbf{S} = \mathbb{E}(\mathbf{g}_i\mathbf{g}_i')$ under Assumption 2.5.
- This DGP accommodates conditional heteroskedasticity (条件异方差)。
- Assumption 2.5 is stronger than Assumption 2.3.

最小二乘估计的大样本性质

Theorem (asymptotic distribution of the OLS Estimator)

(a) *(Consistency of $\boldsymbol{b}$ for $\boldsymbol{\beta}$) Under Assumptions 2.1-2.4,* $\operatorname{plim}_{n\to\infty} \boldsymbol{b} = \boldsymbol{\beta}$.

(b) *(Asymptotic Normality of $\boldsymbol{b}$) If Assumption 2.3 is strengthened as Assumption 2.5, then*

$$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) \underset{d}{\to} N(\mathbf{0}, \operatorname{Avar}(\boldsymbol{b})) \quad \text{as} \quad n \to \infty, \tag{1}$$

where

$$\operatorname{Avar}(\boldsymbol{b}) = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}.$$

(c) *(Consistent Estimate of $\operatorname{Avar}(\boldsymbol{b})$) Suppose there is available a consistent estimator, $\hat{\mathbf{S}}$, of $\mathbf{S}$. Then, under Assumption 2.2, $\operatorname{Avar}(\boldsymbol{b})$ is consistently estimated by*

$$\widehat{\operatorname{Avar}(\boldsymbol{b})} = \mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1},$$

where

$$\mathbf{S}_{\mathbf{xx}} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \frac{1}{n}\mathbf{X}'\mathbf{X}.$$

Proof.

(a) We first write $\boldsymbol{b}-\boldsymbol{\beta}$ in terms of sample means.

$$\begin{aligned}\boldsymbol{b}-\boldsymbol{\beta} &= \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\cdot\varepsilon_i\right)\\ &\equiv \mathbf{S}_{\mathbf{xx}}^{-1}\bar{\mathbf{g}}.\end{aligned}$$

Since by Assumption 2.2 $\{\mathbf{x}_i\mathbf{x}_i'\}$ is ergodic stationary, $\mathbf{S}_{\mathbf{xx}} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}$. Since $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is invertible by Assumption 2.4, $\mathbf{S}_{\mathbf{xx}}^{-1} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ by Continuous Mapping Theorem. Similarly, $\bar{\mathbf{g}} \underset{p}{\to} \mathbb{E}(\mathbf{g}_i)=\mathbf{0}$. So by Continuous Mapping Theorem again, $\mathbf{S}_{\mathbf{xx}}^{-1}\bar{\mathbf{g}} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{0}=\mathbf{0}$. Therefore, $\operatorname*{plim}_{n\to\infty} \boldsymbol{b}=\boldsymbol{\beta}$.

(b) Write

$$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})=\mathbf{S}_{\mathbf{xx}}^{-1}(\sqrt{n}\bar{\mathbf{g}}).$$

By Assumption 2.5, $\sqrt{n}\bar{\mathbf{g}} \underset{d}{\to} N(\mathbf{0},\mathbf{S})$. So, by Slutsky's Theorem, $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) \underset{d}{\to} N(\mathbf{0},\mathrm{Avar}(\boldsymbol{b}))$.

(c) Since $\mathbf{S}_{\mathbf{xx}} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}$ and $\hat{\mathbf{S}} \underset{p}{\to} \mathbf{S}$, $\widehat{\mathrm{Avar}(\boldsymbol{b})} \underset{p}{\to} \mathrm{Avar}(\boldsymbol{b})$ by Continuous Mapping Theorem. □

s^2 is consistent

Theorem (consistent estimation of error variance)

Let $e_i \equiv y_i - \mathbf{x}_i'\boldsymbol{b}$ be the OLS residual for observation i. Under Assumptions 2.1-2.4,

$$s^2 \equiv \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 \underset{p}{\rightarrow} \mathbb{E}(\varepsilon_i^2),$$

provided $\mathbb{E}(\varepsilon_i^2)$ exists and is finite.

Proof.

Since

$$s^2 = \frac{n}{n-K}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right),$$

Proof.

it suffices to prove that the sample mean of e_i^2 converges in probability to $\mathbb{E}(\varepsilon_i^2)$. The relationship between e_i and ε_i is given by

$$e_i \equiv y_i - \mathbf{x}_i'\boldsymbol{b} = y_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_i'(\boldsymbol{b} - \boldsymbol{\beta}) = \varepsilon_i - \mathbf{x}_i'(\boldsymbol{b} - \boldsymbol{\beta}),$$

so that

$$e_i^2 = \varepsilon_i^2 - 2(\boldsymbol{b} - \boldsymbol{\beta})'\mathbf{x}_i \cdot \varepsilon_i + (\boldsymbol{b} - \boldsymbol{\beta})'\mathbf{x}_i\mathbf{x}_i'(\boldsymbol{b} - \boldsymbol{\beta}).$$

Summing over i, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - 2(\boldsymbol{b} - \boldsymbol{\beta})'\bar{\mathbf{g}} + (\boldsymbol{b} - \boldsymbol{\beta})'\mathbf{S}_{\mathbf{xx}}(\boldsymbol{b} - \boldsymbol{\beta}).$$

Therefore,

$$\text{plim}(\frac{1}{n}\sum_{i=1}^{n} e_i^2) = \text{plim}(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2) = \mathbb{E}(\varepsilon_i^2).$$

□

Estimating S consistently

A consistent estimator of $\mathbf{S} = \mathbb{E}(\mathbf{g}_i\mathbf{g}_i') = \mathbb{E}(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i')$ is required to calculate the estimated asymptotic variance, $\widehat{\mathrm{Avar}(\boldsymbol{b})}$. Let

$$\hat{\mathbf{S}} \equiv \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\mathbf{x}_i\mathbf{x}_i', \tag{2}$$

where $\hat{\varepsilon}_i \equiv y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ is some consistent estimator of $\boldsymbol{\beta}$. For this estimator to be consistent for $\mathbf{S}$, we need to make a fourth-moment assumption about the regressors.

6. Assumption 2.6 (finite fourth moments for regressors): $\mathbb{E}[(x_{ik}x_{ij})^2]$ exists and is finite for all $k, j = 1, 2, \ldots, K$.

Theorem (consistent estimation of $\mathbf{S}$)

Suppose the coefficient estimate $\hat{\boldsymbol{\beta}}$ used for calculating the residual $\hat{\varepsilon}_i$ for $\hat{\mathbf{S}}$ in (2) is consistent, and suppose $\mathbf{S} = \mathbb{E}(\mathbf{g}_i\mathbf{g}_i')$ exists and is finite. Then, under Assumptions 2.1, 2.2, and 2.6, $\hat{\mathbf{S}}$ given in (2) is consistent for $\mathbf{S}$.

Proof.

Following the same steps as in the proof of the previous theorem, we have

$$\hat{\varepsilon}_i \equiv y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}} = y_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \varepsilon_i - \mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

so that

$$\hat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{x}_i \cdot \varepsilon_i + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{x}_i\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Proof.

Therefore, for the (l, m)-th element of $\hat{\mathbf{S}}$,

$$
\begin{aligned}
\hat{\mathbf{S}}_{lm} &\equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_{il} x_{im} \quad (l, m = 1, 2, \ldots, K) \\
&= \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 x_{il} x_{im} - 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \cdot x_{il} x_{im} \varepsilon_i \right) \\
&\quad + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' x_{il} x_{im} \right) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).
\end{aligned}
$$

By Cauchy-Schwarz inequality,[a] and ergodic stationarity, it is easy to show that $\hat{\mathbf{S}}_{lm} \underset{p}{\rightarrow} \mathbf{S}_{lm}$. □

[a] $\mathbb{E}(|f \cdot h|) \leq \sqrt{\mathbb{E}(f^2)\mathbb{E}(h^2)}$.

大样本统计推断

Testing Linear Hypotheses

Theorem (robust t-ratio and Wald statistic)

Suppose Assumptions 2.1-2.5 hold, and suppose there is available a consistent estimate $\hat{\mathbf{S}}$ *of* $\mathbf{S}$*. Then*

(a) *Under the null hypothesis* $\mathbb{H}_0 : \beta_k = \bar{\beta}_k$,

$$t_k \equiv \frac{\sqrt{n}(b_k - \bar{\beta}_k)}{\sqrt{\widehat{\operatorname{Avar}(b_k)}}} = \frac{b_k - \bar{\beta}_k}{SE^*(b_k)} \xrightarrow[d]{} N(0,1),$$

where b_k *is the* k*-th element of* $\boldsymbol{b}$, $\widehat{\operatorname{Avar}(b_k)}$ *is the* (k,k) *element of* $\widehat{\operatorname{Avar}(\boldsymbol{b})}$, *and*

$$SE^*(b_k) \equiv \sqrt{\frac{1}{n}\cdot \widehat{\operatorname{Avar}(b_k)}} = \sqrt{\frac{1}{n}\cdot \left(\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1}\right)_{kk}}.$$

$SE^*(b_k)$ *is called the heteroskedasticity-consistent standard error, (heteroskedasticity-)robust standard error, or White's standard error.*

(b) *Under the null hypothesis* $\mathbb{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, *where* $\mathbf{R}$ *is a* $J \times K$ *matrix of full row rank,*

$$W \equiv n \cdot (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left\{\mathbf{R}[\widehat{\operatorname{Avar}(\boldsymbol{b})}]\mathbf{R}'\right\}^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r}) \xrightarrow[d]{} \chi^2(J).$$

Proof.

(a) It is easy to establish this result using (1) and Slutsky's Theorem.

(b) Write W as

$$W = \mathbf{c}_n' \mathbf{Q}_n^{-1} \mathbf{c}_n \quad \text{where} \quad \mathbf{c}_n \equiv \sqrt{n}(\mathbf{R}\boldsymbol{b} - \mathbf{r}) \text{ and } \mathbf{Q}_n \equiv \mathbf{R}\widehat{\mathrm{Avar}(\boldsymbol{b})}\mathbf{R}'.$$

Under $\mathbb{H}_0$, $\mathbf{c}_n = \mathbf{R}\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta})$. So by (1),

$$\mathbf{c}_n \underset{d}{\to} \mathbf{c} \quad \text{where} \quad \mathbf{c} \sim N(\mathbf{0}, \mathbf{R}\mathrm{Avar}(\boldsymbol{b})\mathbf{R}').$$

Also note that

$$\mathbf{Q}_n \underset{p}{\to} \mathbf{Q} \quad \text{where} \quad \mathbf{Q} \equiv \mathbf{R}\mathrm{Avar}(\boldsymbol{b})\mathbf{R}'.$$

Because $\mathbf{R}$ is of full row rank and $\mathrm{Avar}(\boldsymbol{b})$ is positive definite, $\mathbf{Q}$ is invertible. Therefore,

$$W \underset{d}{\to} \mathbf{c}'\mathbf{Q}^{-1}\mathbf{c}.$$

Since the J-dimensional random vector $\mathbf{c}$ is normally distributed with mean $\mathbf{0}$ and since $\mathbf{Q}$ equals $\mathrm{Var}(\mathbf{c})$,

$$\mathbf{c}'\mathbf{Q}^{-1}\mathbf{c} \sim \chi^2(J).$$

□

Testing Nonlinear Hypotheses

Theorem (continued)

(c) *Under the null hypothesis with J restrictions $\mathbb{H}_0 : \mathbf{a}(\boldsymbol{\beta}) = \mathbf{0}$ such that $\mathbf{A}(\boldsymbol{\beta})$, the $J \times K$ matrix of continuous first derivatives of $\mathbf{a}(\boldsymbol{\beta})$, is of full row rank, we have*

$$W \equiv n \cdot \mathbf{a}(\boldsymbol{b})' \left\{ \mathbf{A}(\boldsymbol{b}) \widehat{\mathrm{Avar}(\boldsymbol{b})} \mathbf{A}(\boldsymbol{b})' \right\}^{-1} \mathbf{a}(\boldsymbol{b}) \underset{d}{\to} \chi^2(J).$$

Proof.

(c) The asymptotic result given in (1) and Delta Method imply that

$$\sqrt{n}[\mathbf{a}(\boldsymbol{b}) - \mathbf{a}(\boldsymbol{\beta})] \underset{d}{\to} \mathbf{c}, \quad \mathbf{c} \sim N(\mathbf{0}, \mathbf{A}(\boldsymbol{\beta}) \mathrm{Avar}(\boldsymbol{b}) \mathbf{A}(\boldsymbol{\beta})').$$

Under $\mathbb{H}_0$, $\mathbf{a}(\boldsymbol{\beta}) = \mathbf{0}$, therefore

$$\sqrt{n}\mathbf{a}(\boldsymbol{b}) \underset{d}{\to} \mathbf{c}, \quad \mathbf{c} \sim N(\mathbf{0}, \mathbf{A}(\boldsymbol{\beta}) \mathrm{Avar}(\boldsymbol{b}) \mathbf{A}(\boldsymbol{\beta})').$$

Proof.

Since $\boldsymbol{b} \underset{p}{\rightarrow} \boldsymbol{\beta}$, by Continuous Mapping Theorem, $\mathbf{A}(\boldsymbol{b}) \underset{p}{\rightarrow} \mathbf{A}(\boldsymbol{\beta})$. Hence,

$$\mathbf{A}(\boldsymbol{b})\widehat{\mathrm{Avar}(\boldsymbol{b})}\mathbf{A}(\boldsymbol{b})' \underset{p}{\rightarrow} \mathbf{A}(\boldsymbol{\beta})\mathrm{Avar}(\boldsymbol{b})\mathbf{A}(\boldsymbol{\beta})' = \mathrm{Var}(\mathbf{c}).$$

Because $\mathbf{A}(\boldsymbol{\beta})$ is of full row rank and $\mathrm{Avar}(\boldsymbol{b})$ is positive definite, $\mathrm{Var}(\mathbf{c})$ is invertible. Then

$$\sqrt{n}\mathbf{a}(\boldsymbol{b})' \left\{ \mathbf{A}(\boldsymbol{b})\widehat{\mathrm{Avar}(\boldsymbol{b})}\mathbf{A}(\boldsymbol{b})' \right\}^{-1} \sqrt{n}\mathbf{a}(\boldsymbol{b}) \underset{d}{\rightarrow} \mathbf{c}'\mathrm{Var}(\mathbf{c})^{-1}\mathbf{c} \sim \chi^2(J).$$

□

Implications of Conditional Homoskedasticity

7. Assumption 2.7 (conditional homoskedasticity): $\mathbb{E}(\varepsilon_i^2|\mathbf{x}_i) = \sigma^2 > 0$.

Theorem (large-sample properties of $\boldsymbol{b}$, t, and F under Conditional Homoskedasticity)

Suppose Assumptions 2.1-2.5 and 2.7 are satisfied. Then

(a) *(Asymptotic distribution of **b**) The OLS estimate **b** of $\boldsymbol{\beta}$ is consistent and asymptotically normal with $\mathrm{Avar}(\boldsymbol{b}) = \sigma^2\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$.*

(b) *(Consistent estimation of asymptotic variance) Under the same set of assumptions, $\mathrm{Avar}(\boldsymbol{b})$ is consistently estimated by $\widehat{\mathrm{Avar}(\boldsymbol{b})} = s^2\mathbf{S}_{\mathbf{xx}}^{-1} = n \cdot s^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$.*

(c) *(Asymptotic distribution of the t and F statistics of the finite-sample theory) Under $\mathbb{H}_0 : \beta_k = \bar{\beta}_k$, the usual t-ratio is asymptotically distributed as $N(0,1)$. Under $\mathbb{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, $J \cdot F$ is asymptotically $\chi^2(J)$, where F is the F statistic and J is the number of restrictions in $\mathbb{H}_0$.*

Proof.

(a) Under Assumption 2.7, we have

$$\mathbf{S} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_i'\varepsilon_i^2) = \mathbb{E}[\mathbf{x}_i\mathbf{x}_i'\mathbb{E}(\varepsilon_i^2|\mathbf{x}_i)] = \sigma^2\boldsymbol{\Sigma}_{\mathbf{xx}}.$$

Therefore, $\mathrm{Avar}(\boldsymbol{b}) = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} = \sigma^2\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$.
(b) By ergodic stationary, $\mathbf{S}_{\mathbf{xx}} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}$. Therefore, $\mathbf{S}_{\mathbf{xx}}^{-1} \underset{p}{\to} \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$. Note that s^2 is consistent for σ^2 under Assumptions 2.1-2.4. So, $s^2\mathbf{S}_{\mathbf{xx}}^{-1} \underset{p}{\to} \sigma^2\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$.[a]
(c) Recall that the usual t-ratio is given by

$$t_k \equiv \frac{b_k - \bar{\beta}_k}{SE(b_k)} \equiv \frac{b_k - \bar{\beta}_k}{\sqrt{s^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}}}.$$

And, the robust t-ratio is

$$t_k = \frac{b_k - \bar{\beta}_k}{SE^*(b_k)} \quad \text{with} \quad SE^*(b_k) = \sqrt{\frac{1}{n} \cdot \widehat{\mathrm{Avar}(b_k)}} = \sqrt{s^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}}.$$

So the usual finite-sample t-ratio is numerically identical to the robust t-ratio, and hence asymptotically distributed as $N(0,1)$.

---

[a] Under conditional homoskedasticity, we do not need Assumption 2.6 for consistency.

Proof.

(c) Similarly,

$$
\begin{aligned}
W &= n \cdot (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left\{ \mathbf{R}[\widehat{\mathrm{Avar}(\boldsymbol{b})}]\mathbf{R}' \right\}^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r}) \\
&= n \cdot (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left\{ \mathbf{R}[n \cdot s^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}' \right\}^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r}) \\
&= (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left\{ \mathbf{R}[(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}' \right\}^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r}) / s^2 \\
&= J \cdot F \\
&= (SSR_R - SSR_U)/s^2.
\end{aligned}
$$

Thus, the Wald statistic W is numerically identical to $J \cdot F$. □

特例：解释变量联合显著性检验

For a regression with a constant, consider the null hypothesis that the coefficients of the $K-1$ nonconstant regressors are all zero. We have shown that, under conditional homoskedasticity,

$$F = \frac{R^2/(K-1)}{(1-R^2)/(n-K)}.$$

From this equation, we can derive:

$$nR^2 = \frac{1}{\dfrac{n-K}{n} + \dfrac{1}{n}(K-1)F}(K-1)F.$$

Since $(K-1)F$ converges in distribution to a random variable, $\dfrac{1}{n}(K-1)F \underset{p}{\to} 0$. Then, the asymptotic distribution of the RHS is the same as that of $(K-1)F$, which is $\chi^2(K-1)$.

Resampling Methods: Jackknife and Bootstrap

Jackknife

- The jackknife estimates moments of estimators using the distribution of the leave-one-out estimators.
- Let $\hat{\boldsymbol{\theta}}$ be any estimator of a vector-valued parameter $\boldsymbol{\theta}$ which is a function of a random sample of size n. Let $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} = \operatorname{Var}(\hat{\boldsymbol{\theta}})$ be the variance of $\hat{\boldsymbol{\theta}}$.
- Define the leave-one-out estimators $\hat{\boldsymbol{\theta}}_{(-i)}$ which are computed using the formula for $\hat{\boldsymbol{\theta}}$ except that observation i is deleted.
- The jackknife estimator for $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}}$ is defined as

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}^{\text{jack}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}} \right)', \tag{3}$$

where $\bar{\boldsymbol{\theta}}$ is the sample mean of the leave-one-out estimators

$$\bar{\boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_{(-i)}.$$

To motivate (3), consider the case where $\hat{\theta}$ is the mean of $(x_1, \ldots, x_n)$. We have

$$
\begin{aligned}
\hat{\theta}_{(-i)} &= \frac{1}{n-1}\sum_{j\neq i} x_j = \frac{n}{n-1}\bar{x} - \frac{1}{n-1}x_i, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \\
\bar{\theta} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(-i)} = \bar{x}, \\
\hat{\theta}_{(-i)} - \bar{\theta} &= \frac{1}{n-1}(\bar{x} - x_i).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\hat{V}_{\hat{\theta}}^{\text{jack}} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\right)^2(\bar{x}-x_i)^2 \\
&= \frac{1}{n}\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(\bar{x}-x_i)^2.
\end{aligned}
$$

This is the conventional estimator for the variance of $\bar{x}$.

- Formula (3) is quite general and does not require any technical calculations. However, it requires n separate estimations, which in some cases can be computationally costly.
- In most cases $\hat{V}_{\hat{\boldsymbol{\theta}}}^{\text{jack}}$ will be similar to a robust asymptotic variance matrix estimator.
- The main attractions of the jackknife estimator are that it can be used when an explicit asymptotic variance formula is not available.

Bootstrap

- The bootstrap distribution is obtained by estimation on independent samples created by i.i.d. sampling (sampling with replacement) from the original dataset.
- Let B be the number of bootstrap draws. The bootstrap estimator of variance of an estimator $\hat{\boldsymbol{\theta}}$ is the sample variance across the bootstrap draws:

$$\begin{aligned}\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}^{\text{boot}} &= \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\boldsymbol{\theta}}^*(b)-\bar{\boldsymbol{\theta}}^*\right)\left(\hat{\boldsymbol{\theta}}^*(b)-\bar{\boldsymbol{\theta}}^*\right)', \\ \bar{\boldsymbol{\theta}}^* &= \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{\theta}}^*(b),\end{aligned}$$

where $\hat{\boldsymbol{\theta}}^*(b)$ denotes the bootstrap estimate of $\boldsymbol{\theta}$ based on the b-th bootstrap sample.

- Let $\left\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\right\}$ be an i.i.d. sample of bootstrap estimates of a parameter θ.
- For any $0 < \tau < 1$, one can calculate the empirical quantile q_τ^* of these bootstrap estimates.
- The percentile bootstrap $100(1-\tau)\%$ confidence interval is

$$C^{\text{pc}} = \left[q_{\tau/2}^*, \quad q_{1-\tau/2}^*\right].$$

For example, if $B = 1000$, $\tau = 0.05$, and the empirical quantile estimator is used, then $C^{\text{pc}} = \left[\hat{\theta}_{(25)}^*, \quad \hat{\theta}_{(975)}^*\right]$.

Bootstrap Hypothesis Tests

To test $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$, the bootstrap t-statistic is

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{s(\hat{\theta}^*)},$$

where $\hat{\theta}^*$ is the bootstrap estimator of θ and $s(\hat{\theta}^*)$ is the bootstrap standard error. The bootstrap p-value is given by

$$p^* = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(|t^*(b)| > |t|),$$

that is, the percentage of bootstrap t-statistics that are larger than the observed t-statistic.

When standard errors are not available or are not reliable, we can use the non-studentized statistic $t = \hat{\theta} - \theta_0$. The bootstrap version is $t^* = \hat{\theta}^* - \hat{\theta}$. The bootstrap p-value is

$$p^* = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1} \left(\left| \hat{\theta}^*(b) - \hat{\theta} \right| > \left| \hat{\theta} - \theta_0 \right| \right).$$

Similarly, the bootstrap Wald statistic for testing $\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is

$$W^* = \left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} \right)' \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}^{*-1} \left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} \right).$$

The bootstrap p-value is given by

$$p^* = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1} \left(W^*(b) > W \right).$$

If a reliable covariance matrix estimator $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}$ is not available, a Wald-type test can be implemented with any positive-definite weight matrix instead of $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}$.

Bootstrap for Regression

Consider the linear regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i|\mathbf{x}_i) = 0, \quad \mathbb{E}(\varepsilon_i\varepsilon_j) = 0, \forall i \neq j$$

where there are n observations.

1. The Residual Bootstrap
   - Obtain the OLS estimates $\boldsymbol{b}$ and residuals e_i.
   - Obtain the rescaled residuals using $u_i \equiv \left(\frac{n}{n-K}\right)^{1/2} e_i$.
   - Generate the bootstrap sample by the equation

$$y_i^* = \mathbf{x}_i'\boldsymbol{b} + u_i^*, \quad u_i^* \sim \text{EDF}(u_i).$$

Remark: The residual bootstrap is not valid if the error terms are not independently and identically distributed.

2 The Wild Bootstrap

- Obtain the OLS estimates $\boldsymbol{b}$ and residuals e_i.
- Generate the bootstrap sample by the equation

$$y_i^* = \mathbf{x}_i' \boldsymbol{b} + e_i v_i^*,$$

where v_i^* is a random variable with mean 0 and variance 1. In practice, the following two-point distribution is most commonly used to generate v_i^*:

$$v_i^* = \begin{cases} \frac{-(\sqrt{5}-1)}{2} & \text{with probability } \frac{(\sqrt{5}+1)}{2\sqrt{5}} \\ \frac{(\sqrt{5}+1)}{2} & \text{with probability } \frac{(\sqrt{5}-1)}{2\sqrt{5}}. \end{cases}$$

References

Fumio Hayashi. (2000)

Econometrics, Chapter 2.

Princeton University Press, 2000.

Jeffrey M. Wooldridge. (2010)

Econometric Analysis of Cross Section and Panel Data, 2nd, Chapter 3.

The MIT Press, 2010.

Bruce E. Hansen. (2019)

Econometrics, Chapter 10.

https://www.ssc.wisc.edu/ bhansen/econometrics/ Link

James G. MacKinnon. (2006)

Bootstrap Methods in Econometrics.

The Economic Record **82**: S2–S18.